1. (5 marks)

Listed below are the numbers of deaths from lighning strikes in the United States each year for a sequence of 14 recent and consecutive years.

51 44 51 43 32 38 48 45 27 34 29 26 28 23

* 1. Assuming the population of number of deaths from lightning foolws a normal distribution, construct a 99% confidence interval for population mean and interpret.
  2. Use the data to with a 0.01 significance level to test the claim that the mean number of annual deaths is less than the mean of 72.6 deaths recorded in 1980s.

1. (10 marks)

Data Set 14 “Passive and Active Smoke” includes cotinine levels measured in a group of nonsmokers exposed to tobacco smoke and a group of non smokers not exposed to tobaco smoke. Cotinin is a metabolic of nicotine, meaning that when nicotine is absorbed by the body, cotinine is produced.

(Data description: The data are from 120 subjects in three groups, Smokes, includes subjects who are sokers, ETS includes nonsmokers exposed to environmental tobacco smoke, and NOETS includes nonsmokers not exposed to environmental tobacco smoke. All values are measured leels of serum cotinine (in ng/mL). Dtaa are from US department of Health and Human Services, National Center for Health Statistics, Third national Health and Nutrition Examination Survey.) (a) Use a 0.05 significane level to test th claim that the nonsmokers exposed to tobacco smoke have a higher mean of cotinine level thn nonsmokers not exposed to tobacco smoke.

(b) Construct a confidence interval appropriate for the hypothesis test in part (a).

1. (5 marks)

A study of survival in a paediatric population followed patients a given bioscaffold to treat their heart condition.

|  |  |  |
| --- | --- | --- |
|  |  | Survival  Yes No |
| Diagnosis | VSD  Other | 18 0  5 2 |

Is a VSD diagnosis associated with greater survival rate?

1. (10 marks)

Consider the data on the length of right feet in women in the file “footlength.csv” on the LMS.

* 1. Draw a simple random sample of size 20 from the population (assume you have data for the entire population) and test whether or not the average foot length is 25.5cm. (HINT: use the sample command in R).
  2. Repeat part (a). Did you reach the same conclusion? Explain why you think this was the case.
  3. It turns out that the researchers forgot to mention that the data were actually from two different populations - group one members are professional dancers, and group two members are yoga instructors. The file “footlength2.csv” contains this additional information. Draw a simple random sample of size 20 from each population. Do your samples provide sufficient evidence for you to conclude that women who are professional dancers have smaller feet than their yoga instructor counterparts?

1. (15 marks)

The Obesity data on the LMS contains the BMI (defined as weight (in kg) divided by height (in metres squared)) and waist circumference (in cm) for 230 people, collected as part of a longitudinal cohort. Both measures are used to help identify if someone is overweight, or at risk for chronic disease.

* 1. Produce a scatterplot and describe the relationship between the two measurements.
  2. Carry out a regression analysis which predicts BMI from waist circumference. Test to see if there is any apparent linear relationship between these variables. Your analysis should include appropriate scatter-plot, residual plot and statistical analysis with conclusions.
  3. A doctor needs to know the average BMI of an adult with waist circumferences of 100 cm. Produce a 95% confidence interval for your answer. What if the waist measurement is 150cm instead of 100 cm?
  4. Suppose now an adult with a waist circumference of 100cm wants to predict their own BMI, what would it be? Obtain a 95% prediction interval for your answer.
  5. What would you predict the waist circumference of an individual to be, if their BMI was 30?